

CODIFEXA – Colecciones Digitales para Exactas

Prog. Luis Alberto Olguin; Mag. Raúl Oscar Klenzi

Instituto de Informática – Departamento de Informática

Facultad de Ciencias Exactas, Físicas y Naturales

Universidad Nacional de San Juan

lolguin@iinfo.unsj.edu.ar; rauloscarklenzi@gmail.com

Resumen

La construcción de repositorios institucionales (RI) por parte de facultades de universidades argentinas es cada vez más significativa, pero para que este tipo de emprendimientos – generalmente con presupuestos ajustados – tenga sustentabilidad en el tiempo es necesario “mostrar que es útil” para la institución - los docentes en particular - en el sentido que permite aumentar la visibilidad de la producción científica local y por ende el impacto entre sus pares.

El proyecto que aquí describimos, Colecciones Digitales para la Facultad de Exactas (CODIFEXA), es el resultado del trabajo interdisciplinario de informáticos y bibliotecarios cuyo objetivo es sentar las bases para construir un RI para la Facultad de Ciencias Exactas de la Universidad Nacional de San Juan siguiendo la premisa de que en este tipo de emprendimientos “el centro de atención no es la tecnología, sino en las actividades humanas que se sirven de la tecnología”.

Contexto

Este trabajo se enmarca en las actividades previstas en los proyectos “Minería de Datos en la determinación de perfiles de uso y perfiles de usuarios” (Código 21/E889) y “Colecciones Digitales para la Facultad de Exactas” (Código 21/E903), ejecutados por el Instituto de Informática y el Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de San Juan.

Estos proyectos fueron aprobados durante la convocatoria para la ejecución de proyectos bianuales realizada por CICITCA-UNSJ (2011-2012).

Introducción

El concepto de objeto digital está presente en una abundante bibliografía. Testa [1] indica que se trata de un conjunto de “información nacida en soportes físicos que ha sido digitalizada y también documentos nacidos en formato digital”.

El conjunto de estos objetos digitales, generados por una misma institución y organizados de manera eficiente, conforma un repositorio institucional (RI), cuyo objetivo es la mejora en las posibilidades de encontrabilidad, recuperación y lectura por parte de otros.

Comenzar con la implementación de un RI no es tarea trivial. Murray^[2] indica la importancia de entender que es “a system of people, practices, values and technologies in a particular local environment”.

El “sistema de personas” lo conforman autores, estudiantes, bibliotecarios, lectores y “web surfers”. Las “prácticas” se enmarcan en las acciones de auto-archivo, publicación, difusión, almacenamiento, leer, descargar. Los “valores” son los que se asocian al libre acceso a la información, la comunicación académica, la difusión de la investigación. Por último, las “tecnologías” se asocian al software a utilizar en el RI, los formatos de archivos, metadatos aplicados, la posibilidad de “cosecha” por parte de otros repositorios.

Según cita Tramullas^[3], los RI “se van convirtiendo progresivamente en los lugares en los que numerosas organizaciones almacenan y difunden el resultado de sus actividades. La importancia de este tipo de recursos de información comienza a considerarse como estratégica”.

Al igual que en muchos casos en Argentina, se decidió el estudio de software libre para la implementación de RI en lugar del desarrollo de una solución propia. En la actualidad ya no se piensa que estos software son desarrollos específicos para cada unidad sino que por el contrario se sustentan en equipos de desarrollo fuertes, sostenidos por instituciones o proyectos que permiten asegurar el financiamiento necesario para

asegurar la mejora e innovación del producto a lo largo del tiempo. Una de las características destacables de la adopción de software libre es la adaptabilidad, lo que permite aplicar soluciones locales al momento de la implementación.

Uno de estos softwares es Greenstone, incubado dentro del Proyecto Biblioteca Digital de Nueva Zelanda, Universidad de Waikato y distribuido en colaboración con la UNESCO y la FAO con sede en Amberes, Bélgica.

Nuestro proyecto, Colecciones Digitales para Exactas – en adelante CODIFEXA – adopta este software como plataforma para la implementación del repositorio fundamentalmente por sus facilidades de operación, lo que permite que rápidamente los bibliotecarios hagan uso de su interfaz, y porque responde al concepto de “Acceso abierto, Código abierto y Estándares abiertos”.

En este sentido, Corrado^[4] afirma que las instituciones se beneficiarán trabajando con estos tres componentes en conjunto ya que es posible reducir costos de licencias de software, garantizar el acceso (abierto) - a largo plazo - a las publicaciones generadas por la institución sin las trabas que generalmente se presentan en las publicaciones por suscripción y continuar trabajando con estándares para la descripción de los recursos, lo que garantiza la interoperabilidad, visibilidad y encontrabilidad de la literatura.

Fournier^[5] al realizar una descripción de metadatos indica que “entre la gran

variedad de formatos de metadatos existentes en Internet, la iniciativa Dublin Core [6] es la más citada y aceptada; es un formato producto de un esfuerzo internacional e interdisciplinario con una vida muy intensa y el más influyente en relación con el desarrollo de la teoría del uso de los metadatos para la recuperación de información en la red”.

En nuestra investigación en pos de definir el conjunto de metadatos aplicar en el RI, encontramos que muchos proyectos exitosos aplican Dublin Core, tal es el caso de la Biblioteca Digital Colombiana en la que confluyen 73 repositorios con más de 80.000 documentos digitales.

Así mismo desde el Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina, se publicó la recomendación de adopción de este modelo de metadatos para participar en el Sistema Nacional de Repositorios.

Actualmente CODIFEXA aplica el conjunto Dublin Core Simplificado, generando los metadatos a partir de los registros catalográficos creados por la biblioteca de la facultad de Exactas. Para esta tarea se diseñaron herramientas de software que permiten extraer la información de los campos de la base de datos CDS/Isis hacia archivos DC-XML que luego son importados por Greenstone.

El protocolo OAI-PMH es el estándar abierto fundamental para asegurar la exposición, agregación, acceso e interoperabilidad de los contenidos depositados en los repositorios. El

protocolo OAI-PMH es muy sencillo y requiere el uso de otros estándares y protocolos, como el protocolo HTTP (protocolo Web) y el estándar Dublin Core para el formato de los metadatos que permite la interoperabilidad entre los repositorios.

Greenstone, desde su versión 2.84, posee un “OAI data provider” lo que asegura que el contenido del repositorio podrá ser cosechado por otros, aumentando la visibilidad del mismo.

Un RI sin la participación de sus actores principales, los autores de los trabajos, es imposible que logre apoyo institucional en el tiempo. Keefer^[6] destaca que “los autores tienen mucho por ganar porque la presencia de sus trabajos en los repositorios aumentará la visibilidad y estabilidad, lo que presumiblemente contribuirá a recibir un mayor número de citas”.

Consientes que uno de los principales obstáculos para introducir documentos es el “miedo al plagio”, se conformó un grupo de trabajo en el proyecto que se encarga de las pautas de difusión y el estudio de normativas de otras universidades respecto a la cesión de derechos de publicación.

La difusión de los beneficios de contar con RI en la UNSJ es labor de este grupo, habiéndose elaborado una encuesta, que se encuentra actualmente vigente, para consultar a los docentes/investigadores acerca de sus actitudes frente al depósito de los trabajos por ellos generados. Resultados parciales muestran que un

26% de los docentes de la facultad han generado más de 10 artículos científicos en los últimos años; el área que más investigación muestra es Informática con un 43%; el 71% expresa desconocer el concepto “archivo institucional” y un 18% manifiesta que difunde sus publicaciones en sitios externos (Blog, Wikis, etc.).

El recurso humano de CODIFEXA conforma, como lo define Hernandez Arias^[7], un grupo de trabajo en el cual “un conjunto de individuos realizan actividades con la finalidad de lograr un objetivo específico”. Dado que los integrantes del proyecto no comparten el mismo espacio de trabajo, el planteo de trabajo colaborativo se basó en un mix de trabajo presencial y telemático. Con el fin de aplicar este último, se realizó el estudio de plataformas “Site As A Service” habiéndose seleccionado TeamLab como plataforma de actividad en el proyecto. Dentro de este espacio virtual de trabajo se comunican, coordinan, discuten, etc. las tareas asignadas a cada participante.

“No solo se trata de la interconexión en red de la tecnología sino la interconexión en red de los seres humanos mediante la tecnología”. Para nuestro equipo de trabajo, en particular para el grupo de bibliotecarios, aplicar tecnología en el modelo colaborativo ha sido novedoso y ha permitido organizar y optimizar el tiempo productivo.

Líneas de investigación y desarrollo

Dentro del proyecto, que se ajusta a una investigación de informática aplicada, la línea principal es la aplicación de soluciones open source para el diseño e implementación de repositorios institucionales.

Las acciones llevadas adelante abarcan temas referidos a la implementación de software libre en diferentes plataformas, el estudio y aplicación de protocolos de interoperabilidad en repositorios, el manejo de grandes volúmenes de información mediante la aplicación de la API Lucene para recuperación en texto completo.

Un área de especial interés que surge como línea de investigación en el proyecto es la minería web, en particular lo referido a la aplicación de soluciones de código abierto (o gratuito) para obtener métricas de accesibilidad a los contenidos de repositorios. Este trabajo se lleva adelante en forma conjunta con el grupo de investigación del proyecto “Minería de Datos en la determinación de perfiles de uso y perfiles de usuarios” y apunta a formar recursos humanos capaces de interactuar con estas herramientas para aplicarlas en futuras actualizaciones del repositorio en construcción.

También forma parte de las líneas de investigación abordadas por el proyecto la preservación de objetos digitales, tarea que se encara en forma conjunta con las actividades que desarrolla el equipo del proyecto “*Red COES de repositorios de*

acceso abierto - PICT-O 2010-132” y tiene como objetivo realizar un relevamiento de las actuales técnicas de preservación aplicadas a nivel internacional a fin de realizar recomendaciones en esta área para ser aplicadas al repositorios en creación.

Resultados y Objetivos

Dentro de los objetivos planteados en CODIFEXA se han alcanzado los referidos a “implementar un repositorio institucional con el objeto de almacenar la producción intelectual generado por la Facultad de Ciencias Exactas de la Universidad Nacional de San Juan”. Actualmente se encuentran publicadas 70 tesis de licenciaturas en Informática, Geología y Geofísica.

El software aplicado es Greenstone versión 2.85 sobre el que se trabajó aplicando HTML5 y CSS3 con el objeto de personalizar la interfaz de usuario. El motor de búsqueda aplicado es Lucene, provisto por Greenstone.

La tabla “crosswalk” de migración de registros desde formato LIBRI-UNSJ hacia XML-DC y el software asociado forma parte de los objetivos alcanzados por el proyecto.

Se pretende alcanzar nuevos objetivos relacionados al Open Access y los autores, a partir de los resultados que se obtengan al completar el estudio de respuestas a la encuesta lanzada a docentes de la FCEFN.

Formación de RRHH

Durante este periodo integrantes del proyecto han trabajado en temas de investigación conjuntos con los proyectos “Minería de Datos en la Determinación de Patrones de Uso y Perfiles de Usuarios” (cód. 21/ E889) y “Red COES de Repositorios de Acceso Abierto” (PICT-O CIN 2010-Agencia Nacional de Promoción Científica y Tecnológica).

La formación de los integrantes del proyecto se concretó en la participación de cursos acerca de aplicación de metadatos, construcción de colecciones digitales, open Access.

La transferencia al medio se realizó mediante el dictado de charlas, talleres referidos a los temas de investigación abordados.

Actualmente se encuentran en curso la concreción de tesis de grado de alumnos de la Licenciatura en Informática que abordan las temáticas OPAC-Web.

Referencias

[1] TESTA, Patricia. Descripción de Objetos Digitales: METADATOS [en línea]. 2009. Disponible en http://www.bn.gov.ar/descargas/catalogadores/ponencias/251109_09a.pdf

[2] Murray, Annie. What we talk about when we talk about institutional repositories [en línea]. 2008 [fecha de consulta: 5 de marzo 2013]. Disponible en

http://library.concordia.ca/about/staff/forum/What_We_Talk_About_When_%20We_Talk_About.pdf

[3] TRAMULLAS, Jesús. Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones. El profesional de la información. 2006, v. 15, nro. 3 [fecha de consulta: 10 de marzo 2013]. Disponible en http://eprints.rclis.org/9345/1/vol15_3.1.pdf

[4] CORRADO, Edward M. The Importance of Open Access, Open Source, and Open Standards for Libraries [en línea]. 2005. [fecha de consulta: 7 de marzo 2013]. Disponible en <http://www.istl.org/05-spring/article2.html>

[5] FOUNIER, Isabel. Descripción de los recursos de información en Internet: formato Dublín Core [en línea]. ACIMED. 2006, v. 14, nro. 4. [fecha de consulta: 9 de marzo 2013]. Disponible en http://bvs.sld.cu/revistas/aci/vol14_4_06/aci09406.htm

[6] KEEFER, Alice. Los repositorios digitales universitarios y los autores [en línea]. Anales de documentación. 2007, nro. 10. [fecha de consulta: 11 de marzo 2013]. Disponible en <http://revistas.um.es/analesdoc/article/download/1151/1201>

[7] HERNANDEZ ARIAS, Aymara. La tecnología de Trabajo Colaborativo en el contexto universitario [en línea]. Compendium. [fecha de consulta: 11 de

marzo 2013]. Disponible en <http://www.ucla.edu.ve/dac/compendium/compendium6/Tecnologia%20de%20trabajo%20colaborativo.htm>